Human dUTP pyrophosphatase: uracil recognition by a β hairpin and active sites formed by three separate subunits

Clifford D Mol[1], Jonathan M Harris[2], Evan M McIntosh[2] and John A Tainer[1]*

Background: The essential enzyme dUTP pyrophosphatase (dUTPase) is exquisitely specific for dUTP and is critical for the fidelity of DNA replication and repair. dUTPase hydrolyzes dUTP to dUMP and pyrophosphate, simultaneously reducing dUTP levels and providing the dUMP for dTTP biosynthesis. A high cellular dTTP:dUTP ratio is essential to avoid uracil incorporation into DNA, which would lead to strand breaks and cell death. We report the first detailed atomic-resolution structure of a eukaryotic dUTPase, human dUTPase, and complexes with the uracil-containing deoxyribonucleotides, dUMP, dUDP and dUTP.

Results: The crystal structure reveals that each subunit of the dUTPase trimer folds into an eight-stranded jelly-roll β barrel, with the C-terminal β strands interchanged among the subunits. The structure is similar to that of the *E. coli* enzyme, despite low sequence homology between the two enzymes. The nucleotide complexes reveal a simple and elegant way for a β hairpin to recognize specific nucleic acids: uracil is inserted into a distorted antiparallel β hairpin and hydrogen bonds entirely to main-chain atoms. This interaction mimics DNA base pairing, selecting uracil over cytosine and sterically precluding thymine and ribose binding. Residues from the second subunit interact with the phosphate groups and a glycine-rich C-terminal tail of the third subunit caps the substrate-bound active site, causing total complementary enclosure of substrate. To our knowledge, this is the first documented instance of all three subunits of a trimeric enzyme supplying residues that are critical to enzyme function and catalysis.

Conclusions: The dUTPase nucleotide-binding sites incorporate some features of other nucleotide-binding proteins and protein kinases, but seem distinct in sequence and architecture. The novel nucleic acid base recognition motif appears ancient; higher order structures, such as the ribosome, may have evolved from a motif of this kind. These uracil–β-hairpin interactions are an obvious way for peptides to become early coenzymes in an RNA world, providing a plausible link to the protein–DNA world. Within the β hairpin, there is a tyrosine corner motif that normally specifies β-arch connections; this tyrosine motif was apparently recruited to discriminate against ribonucleotides, more recently than the evolution of the β hairpin itself.

Addresses: [1]Department of Molecular Biology, MB-4, The Scripps Research Institute, 10666 North Torrey Pines Rd, La Jolla, CA 92037, USA and [2]Department of Biochemistry, The University of Queensland, Brisbane, Queensland, 4072, Australia.

*Corresponding author.
E-mail: jat@scripps.edu

Key words: crystal structure, DNA repair, enzyme–DNA interactions, nucleotide-recognition motif, protein structure/function, RNA world

Received: 9 June 1996
Revisions requested: 8 July 1996
Revisions received: 30 July 1996
Accepted: 2 August 1996

Structure 15 September 1996, 4:1077–1092

Introduction

DNA integrity and function depends on processes that either exclude or remove the normal RNA base uracil [1]. Uracil can arise in DNA through two distinct pathways: the spontaneous deamination of cytosine, which must be removed by uracil-DNA glycosylase [2], and the utilization of dUTP in place of dTTP by DNA polymerase during DNA replication [3], which must be prevented by dUTP pyrophosphatase (dUTPase; EC 3.6.1.23), also called deoxyuridine triphosphatase or dUTP nucleotidohydrolase). Misincorporation of dUTP by DNA polymerase disrupts specific DNA recognition sequences critical to transcription [4–6] and ultimately defeats the uracil base excision repair pathway. High cellular dUTP concentrations result in many A·U base pairs in DNA, leading to cycles of uracil misincorporation followed by replicative repair, generating multiple DNA-strand breaks and eventual cell death [7–10]. This cycle, termed thymine-less cell death, which is the reason behind targeting thymidylate synthase and dihydrofolate reductase in chemotherapeutic drug design, indicates the key biological role of dUTPase regulation of uracil metabolism.

Thymine-less cell death is precluded because the cellular concentration of dUTP is kept low by dUTPase which catalyzes the breakdown of dUTP to dUMP and PPi. In addition, this mechanism provides the dUMP substrate for dTTP biosynthesis thereby maintaining a high cellular dTTP/dUTP ratio. Thus, in a single reaction dUTPase ensures the fidelity of DNA replication, transcription and uracil base excision repair. Furthermore, dUTPase is essential in both *E. coli* [11] and *Saccharomyces cerevisiae*

[12], which verifies its importance in DNA replication and repair: dUTPase regulation in *Drosophila* suggests a role in programmed cell death [13]. Many animal viruses (including poxviruses, herpesviruses and several vertebrate retroviruses) encode dUTPases and dUTPase-deficient viruses exhibit impaired replication in nondividing cells [14,15]. Both B- and D-type retroviruses, and non-primate lentiviruses encode dUTPases, whereas the HIV Vpr protein binds human uracil-DNA glycosylase [16], so mammalian dUTP metabolism is of broad-based biological importance. Moreover, human dUTPase, along with many other DNA-repair enzymes, is colocalized to the nucleus and mitochondria [17]: the nuclear species is specifically phosphorylated at a serine residue within a consensus phosphorylation site for a cyclin-dependent kinase [18], suggesting that dUTPase is regulated in a cell-cycle-dependent manner.

Inhibitors of human dUTPase may be useful as anticancer chemotherapeutic compounds either alone or in combination with existing treatments [19,20]. The differential effects of inhibitors on the activities of dUTPases from herpes simplex viruses [21] and Epstein-Barr virus [22] versus human dUTPase suggest that a detailed knowledge of the structure of human dUTPase may assist in the design of antiviral agents [23]. The crystal structure of *E. coli* dUTPase showed the enzyme's fold and trimer assembly [24]; however, this trimer was crystallographic, and the essential, sequence-conserved C terminus of each subunit, was not seen in the electron-density maps. Unfortunately, the *E. coli* structure also did not show the location of the enzyme's active site, how the exquisite specificity for dUTP is conferred, or potential mechanisms for cleaving dUTP to dUMP and inorganic pyrophosphate. Thus, detailed structural information pertaining to these key biological issues is lacking, as is the high resolution structural characterization of any eukaryotic dUTPase.

Here, we present the refined crystal structure of trimeric human dUTPase at 2.0 Å resolution, together with structures of dUTPase complexed with bound dUMP, dUDP, and dUTP. The active sites defined in the enzyme–nucleotide structures lie at the three subunit interfaces and comprise conserved residues from all three subunits. Hydrogen-bond pairing between uracil and dUTPase, which occurs by insertion of uracil into a distorted β hairpin of one subunit, represents a new nucleotide-recognition motif with implications for the evolution of RNA–protein and RNA–polypeptide interactions. Nucleotide binding is also notable for ordering the flexible C-terminal tail of the third subunit, which caps the active site and interacts specifically with the bound nucleotides. Structural and mutational analysis (presented herein) implicates conserved residues from each subunit in productive dUTP binding and catalysis, and suggests a possible mechanism for dUTP hydrolysis.

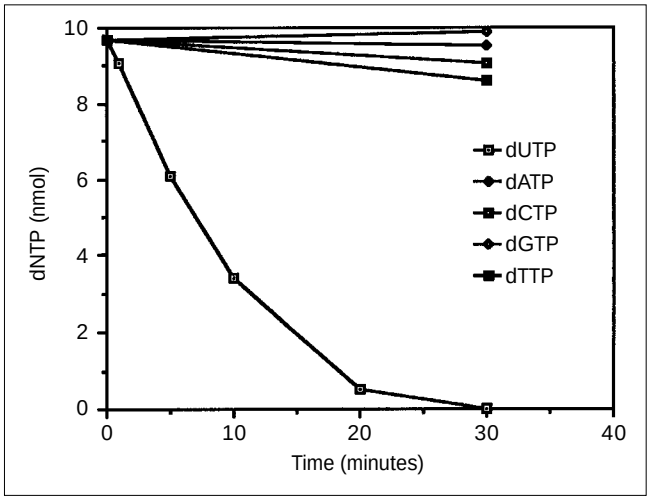
Results and discussion

Activity and specificity of human dUTPase

The dUTPase enzyme that we describe here is a fully active recombinant human enzyme in which the N-terminal 19 amino acids of the nuclear form of the enzyme [17] have been replaced by six histidine residues. This six-His tag was engineered into the cloned gene for human dUTPase [20] to facilitate purification. Details of the cloning, expression in *E. coli*, protein purification and the assay of enzyme activity are given in the Materials and methods section.

Recombinant human dUTPase is exquisitely specific for dUTP. The specific activity of the enzyme, as determined from the average of three independent experiments, is 30.2 ± 2.4 μmol $min^{-1}mg^{-1}$. In contrast, none of the other deoxyribonucleotide triphosphates that are normally found in DNA are hydrolyzed (Fig. 1). In particular, dTTP is not a substrate at the limit of detection of the assay (30 nmol $min^{-1}mg^{-1}$), although a previous report suggested that dTTP was slightly inhibitory to the enzyme and might be a poor substrate [25]. Human dUTPase requires a divalent metal ion for activity, and Mg^{2+} stimulates activity by 165%. Ca^{2+} inhibits and reduces activity to 31%, an effect also seen with dUTPase from rat liver [26]. The pH activity profile for human dUTPase reaches an optimum at pH 8.0, with greater than 90% optimum activity extending over a pH range from 7.0–8.5. Hydrolysis of dUTP is reduced outside this range, with only ~50% enzyme activity at pH 6.0 and ~32% activity at pH 9.0 (data not shown).

Figure 1



The activity of human dUTPase for deoxyribonucleotide triphosphates shows an exquisite specificity for dUTP. Results for dUTP are from a single assay containing 0.2 μg purified enzyme. Assays for the other deoxyribonucleotide triphosphates are an average of two experiments each containing 1.0 μg purified enzyme with a single time point taken at 30 mins. For all assays, control experiments performed with heat-denatured enzyme confirmed that hydrolysis was due to enzyme activity.

Structure determination

Recombinant human dUTPase crystallizes in the orthorhombic space group $P2_12_12$. The three subunits in the crystallographic asymmetric unit form a trimer around a noncrystallographic threefold axis located approximately parallel to the crystallographic twofold axis. The human dUTPase structure was determined by molecular replacement using the *E. coli* dUTPase crystallographic trimer [24] as a search model (see Materials and methods section). Results verified the structural conservation of dUTPase between bacteria and eukaryotes, despite the fact that the *E. coli* and human enzymes share only 31% residue identity (45 from a total of 141 amino acids). The crystal structure determination and refinement statistics are presented in Table 1. In the structure of the uncomplexed enzyme, all of the polypeptide-backbone atoms for residues 1–126, and most of the side-chain atoms for all three subunits are clearly visible in the final $2F_{obs}$-F_{calc} electron-density map, including the region of the water-filled active-site pocket (Fig. 2). The structures of human dUTPase in complex with dUMP, dUDP, and dUTP were determined by difference Fourier techniques between the uncomplexed, human dUTPase structure and the complexes. In these nucleotide-bound dUTPase structures, residues 127–136 become ordered and the corresponding amino acids were placed in the electron density. The bound deoxyribonucleotides are well defined in the electron density from simulated-annealed omit F_{obs}-F_{calc} maps (Fig. 3). Details of the crystallization, data collection, structure solution and refinement of these four structures are given in the Materials and methods section.

Subunit, domain fold and trimer assembly

Each human dUTPase subunit folds into a β barrel with the C-terminal β strand extended outside the subunit domain (Fig. 4a,b) and interchanged among the three subunits, forming the enzymatic trimer (Fig. 4c,d). Each folding domain therefore contains two different polypeptide chains that are tightly interlocked by swapped C-terminal β strands. The domain fold is a distorted eight-stranded β barrel consisting primarily of a six-stranded antiparallel β jelly roll [27], formed in clockwise order by β2, β7, β4, β5, β6, and β3, plus the adjacent and parallel N- and C-terminal strands β1 and β8′, where β8′ is from the adjacent subunit (Fig. 4a–d). This distorted β barrel is closed by a conical cap formed by five antiparallel β strands, which extend from the barrel, plus the single α helix: β2b, β6b, α1, β5, β6, and β3 (Fig. 4a,b). The exchanged C-terminal β strands are also remarkable for continuing outside the domain fold as glycine-rich, flexible tails that cap the substrate-binding sites (Figs 4c,d,5).

Within the β barrel, β5 is broken into two parts by Ala75, which is flanked by two glycines and is the only Ramachandran outlier in the structure. The electron density in this region is clear (Fig. 2) and places Ala75 in the slightly

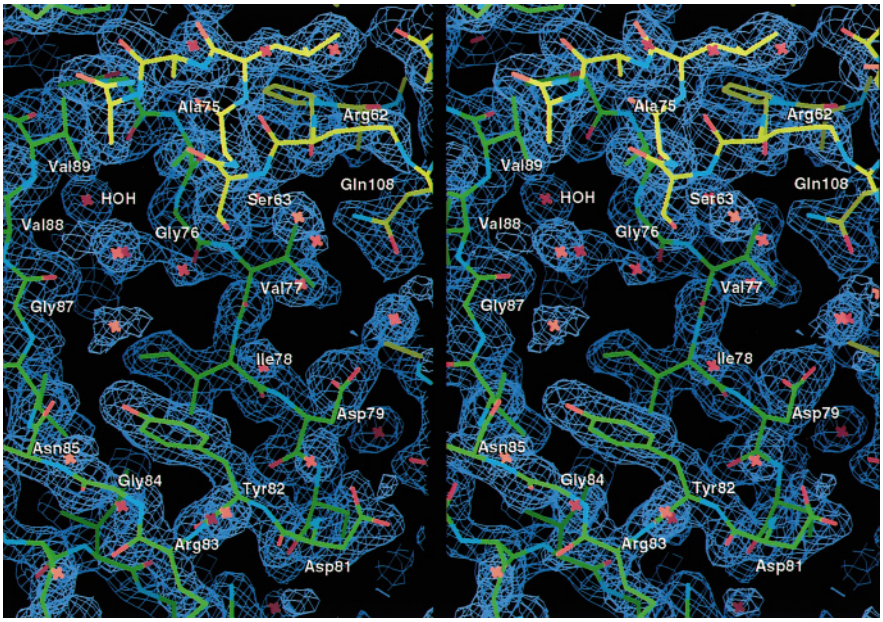
Table 1

Crystallographic data collection and refinement statistics.

Parameters	Native	dUMP	dUDP	dUTP
Unit cell dimensions (a,b,c [Å])	73.0, 110.7, 53.0	71.9, 109.0, 51.2	73.2, 110.8, 53.2	73.4, 111.2, 53.6
Concentration (mM)		50	50	50
Soaking time (h)		18	72	24
Resolution (Å)	2.0	2.1	2.0	1.9
Observations	101 047	78 452	102 902	118 539
Unique reflections	29 007	23 324	28 406	33 600
I/σ	8.3	11.8	10.8	13.3
Completeness (%)	97	96	94	95
R_{sym} (%)*	9.1	8.0	7.7	6.2
Final shell	2.07–2.00	2.18–2.10	2.07–2.00	1.97–1.90
Completeness (%)	99	96	82	73
R_{sym} (%)	45.3	23.5	32.3	29.4
Refinement				
Protein atoms	3049	3231	3231	3231
Ligand atoms		60	72	88
Solvent atoms	288	219	193	188
R_{value} †	0.169	0.188	0.179	0.180
R_{free} ‡	0.270	0.266	0.250	0.238
Rms deviations from ideality				
Bond lengths (Å)	0.017	0.014	0.012	0.014
Bond angles (°)	2.78	1.57	2.34	2.49

*R_{sym} is the unweighted R value on I between symmetry mates.
†$R_{value} = \Sigma_{hkl}|F_{obs}(hkl) - F_{calc}(hkl)|/\Sigma_{hkl}|F_{obs}(hkl)|$. ‡$R_{free}$ = the free R value for 5% reflections against which the model was not refined

Figure 2



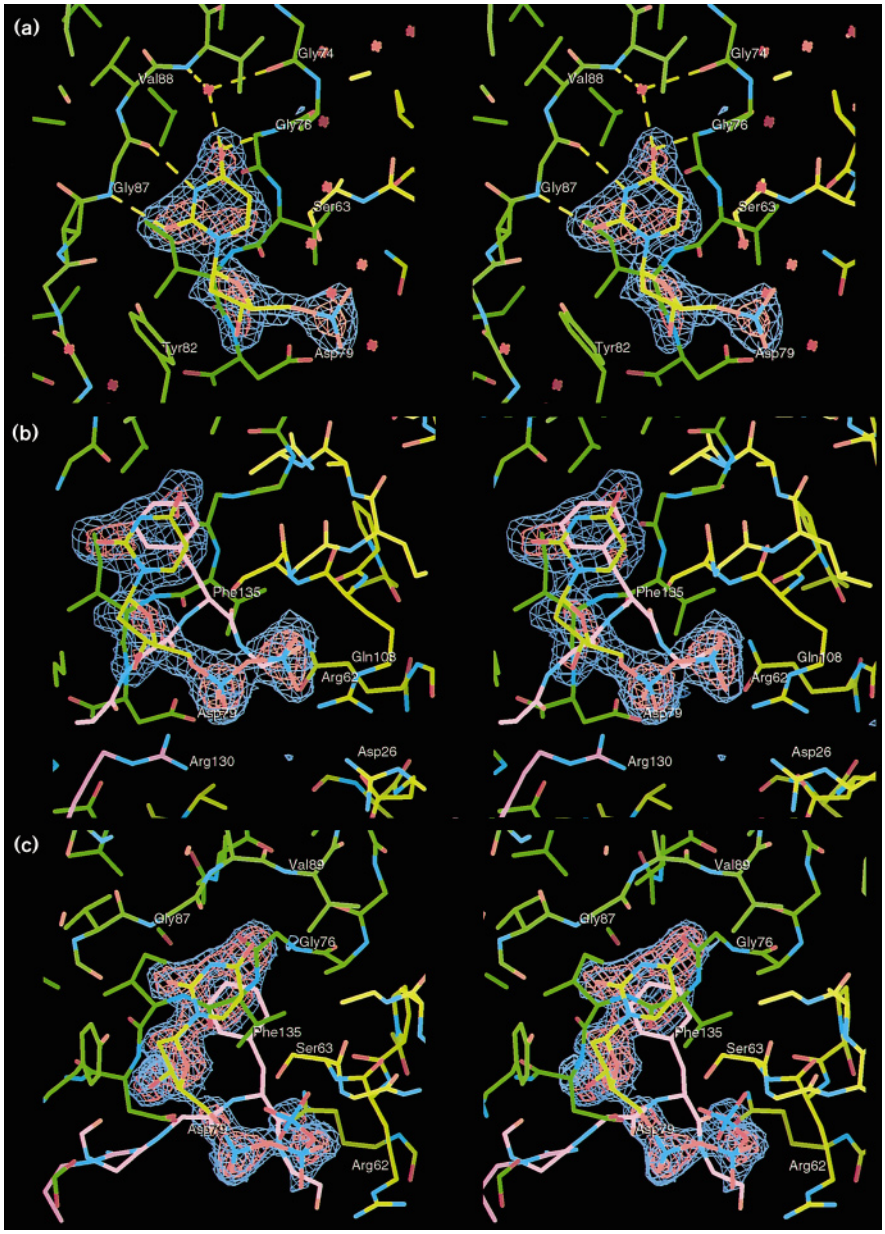
Stereo view of the human dUTPase model and experimental electron density at one of the three active-site pockets, which are each formed by the subunit edge-to-side or lap-joint interfaces. The $2F_{obs}-F_{calc}$ electron-density map (contoured at 1.0 σ) shows the active-site electron density of the uncomplexed enzyme with the atomic coordinates refined at 2.0 Å resolution. The active sites occur at subunit interfaces, with individual subunits distinguished by yellow and green carbon atoms. In the absence of nucleotide substrate, the active-site pocket contains bound water molecules (red crosses), and the C-terminal tail that covers this region when substrate is bound becomes invisible in the electron-density maps. Key residues are labeled as well as the conserved water molecule (HOH) that forms a hydrogen bond to uracil O4 in the nucleotide-bound complexes. (Figure displayed using XFIT [70].)

disfavored γ-turn region of the Ramachandran plot of main-chain torsion angles. The structural effect of this break in β5 is to create a pocket between β5 and β6 that lacks β-strand hydrogen bonds: this pocket is a functionally important feature (Fig. 3 and 'Active-site pocket' section below). Residues 10–24 are extended, lacking regular β-sheet hydrogen bonds except for a hydrogen bond from Thr16 to Asp26. They interact with adjacent strands primarily via side-chain to main-chain interactions. The overall residue distribution (Fig. 5) is typical of β-barrel folds, with the domain interior closely packed with hydrophobic side chains. In cross section, the resulting barrel and conical cap form a triangular-shaped domain with a flat inner face and a sharply curved outer face. There are no disulphide bridges within the subunit fold, although the two cysteine residues of human dUTPase (Cys55 and Cys111) are in close proximity to each other (Cα–Cα distance ~5.8 Å) near the subunit interfaces. These reactive cysteine residues may account for the reported inactivation of human dUTPase by various mercury compounds [21].

The three subunits associate about a threefold axis of symmetry roughly perpendicular to the C-terminal β strands to form an interlocked, propeller-shaped trimer of about 50 Å per side and 40 Å deep (Fig. 4c,d). The inner, flat β sheet of the dUTPase subunit faces the trimer axis. The root mean square (rms) deviation in the Cα positions of these secondary structure elements is 1.20 to 1.28 Å when each of the human dUTPase subunits are compared with the *E. coli* subunit. Thus, the overall fold shows a close resemblance to the *E. coli* enzyme, despite the limited sequence homology (Fig. 5) and the large evolutionary distance between humans and bacteria, suggesting that the dUTPase trimer is strongly conserved in biology. For comparison, the rms deviation in atom positions between the human subunits, both within each trimer and comparing the different free and bound structures, is approximately 0.5 Å for all Cα atoms and between 1.0 and 1.5 Å for all atoms.

A channel runs down the molecular threefold axis, which is open at one end and closed at the other by the three exchanged β8 strands (Fig. 4d). Unlike the hydrophobic channel of the *E. coli* enzyme, the human dUTPase channel contains numerous water molecules, alternating layers of positively and negatively charged amino acids and two potential metal-ion-binding sites. Beginning at the open end of the channel, charged side chains Glu39 and Lys40 are directed into solvent, followed by Asp72, Arg58 and Glu112 side chains directed towards the molecular threefold axis, and Arg113 side chain directed into solvent. The three Arg113 residues form interlocked main-chain hydrogen bonds to close the channel bottom. There is a short hydrogen bond (~2.8 Å) between the Oε2 atom of Glu112 and the Nε atom of Arg58. The three Glu112 Oε1 atoms are directed towards the molecular threefold with very short hydrogen bonds (~2.2 Å) to a probable metal ion, modelled as a water molecule. Similarly, the Asp72 Oδ1 atoms are directed towards a probable metal ion at the molecule threefold axis, with hydrogen-bond lengths of ~2.4 Å. Although these ligands were modelled as water molecules in the refinement so that steric restraints would strongly favor longer distances, the extremely short hydrogens bonds to Asp72 and Glu112 suggest that they may be biologically relevant: they may form metal-ion-binding

Figure 3

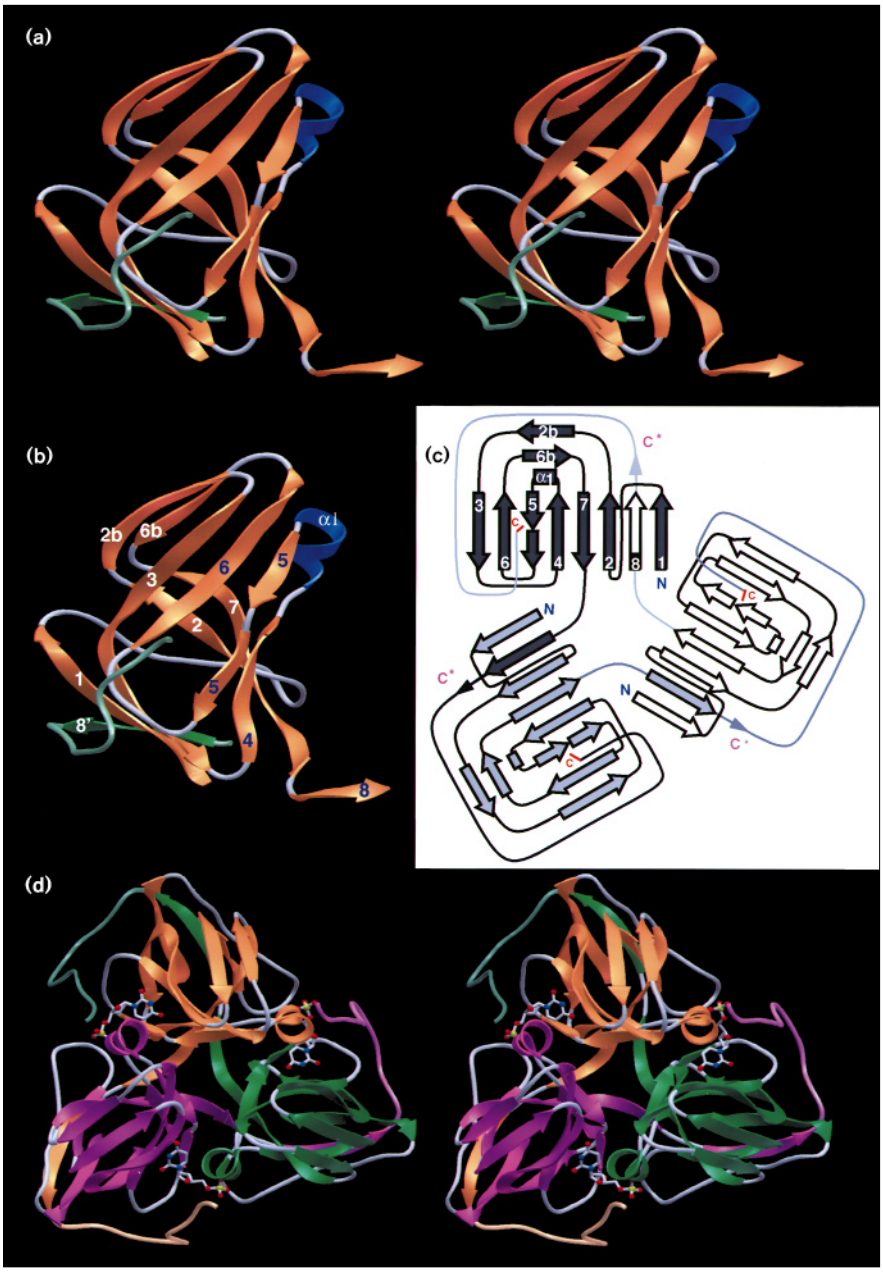


Stereo views of three human dUTPase active-site complexes showing the bound nucleotides. Key residues are labeled and in each case the electron density is from the simulated-annealed $F_{obs}-F_{calc}$ omit map made by excluding the bound nucleotide from the phase calculation. (a) The dUTPase–dUMP complex with electron density, contoured at 3.0 σ (blue) and 6.0 σ (red). Uracil is recognized entirely by main-chain hydrogen bonds (dashed yellow lines) to a β hairpin motif formed by β5–β6. Residues from the two separate subunits are distinguished by yellow and green colored carbon atoms. Positions of ordered solvent molecules are shown by red crosses. (b) The dUTPase–dUDP structure with electron density contoured at 3.0 σ (blue) and 7.0 σ (red). The C-terminal tail residue Phe135 (pink carbons) caps the active-site pocket to exclude solvent. Carbon atoms from the three separate subunits are colored yellow, green and pink. Two subunits contribute arginines for interactions with the substrate phosphate groups (Arg130 and Arg62). Ordered solvent molecules are not shown. (c) The dUTPase–dUTP complex with electron density contoured at 3.0 σ (blue) and 5.0 σ (red). Ser63 and the α1 dipole (lower right) interact with the β-phosphate. Ordered solvent molecules are not shown. (Figure displayed using XFIT [70].)

sites, for example as sites of inhibitory Ca^{2+} binding in mammalian dUTPases [26], or more likely as structural sites for trimer stability.

Due to the polar channel about the molecular threefold, the major subunit associations occur at the pairwise interfaces on the trimer perimeter and with the exchanged C-terminal β strands (Fig. 4d). The sequence-conserved loop between β1 and β2 (Thr16 to Tyr25) is sandwiched between two segments of polypeptide from the adjacent subunit, comprising the turn between β5 and β6 (Asp79 to Gly84), and an extended region of chain between β7 and β8 (Ile114 to Ile119) (Figs 4c,5). Thus, adjacent subunits form broad lap joints (edge-to-side interactions) at the β5 edge. In this same lap-joint interface region, the Phe70 side chain anchors the single α1 helix into the adjacent subunit via Phe91 and Phe93 located on β6. These three residues interact across the threefold to form a phenylalanine cluster adjacent to the putative Asp72 metal-ion-binding site. On the outside edge of the trimer, β8′ (from the adjacent subunit) inserts parallel to β1 (Fig. 4a–c) making extensive van der Waals' and hydrogen-bonding contacts involving residues that are not conserved in the *E. coli* enzyme: Phe115 inserts between Met1 and Arg113, Tyr116 and Pro117 carbonyls hydrogen bond to Arg17 and Tyr25 side chains, and Ile119 takes part in hydrophobic

Figure 4



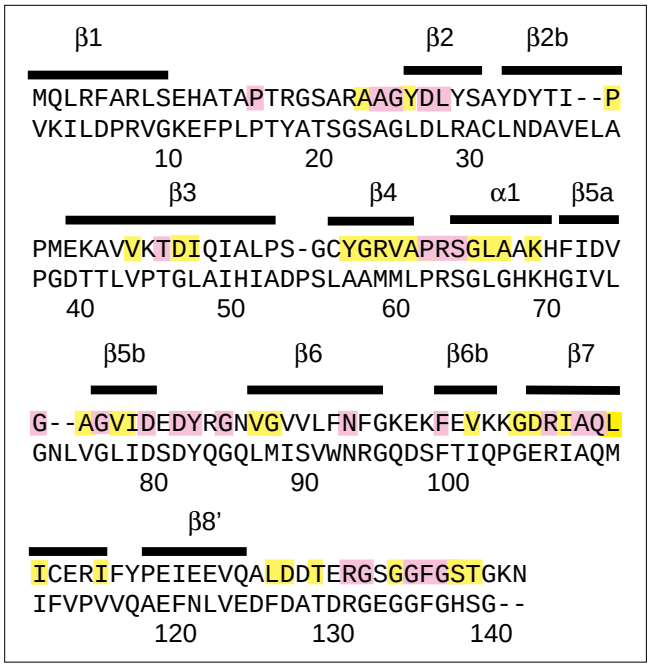
Subunit fold, trimer assembly and active-site location of human dUTPase. **(a)** A stereo ribbon diagram showing the eight-stranded β barrel subunit fold and C-terminal strand exchange. Each triangular domain fold contains the C-terminal strand from a neighboring subunit (green) with the secondary structure within the polypeptide chain color-coded for β strands (orange), α helix (blue), and loop (pale blue). Note the extension of the C-terminal strand away from the domain for insertion into a neighboring domain. **(b)** dUTPase subunit fold with labeled secondary structure in the same orientation and coloring as described in (a). β-strand and α-helix numbering is defined in the text and colored blue and white for clarity. **(c)** Topology diagram for the β strand swapped dUTPase trimer. Each of the three active sites assembles at a subunit interface. The β strands and α helix for one of the subunits are labeled as in (b) and described in the text. C and C* indicate ends of the visualized polypeptide chains for strands swapped among subunits, in the absence and presence of bound nucleotide, respectively. **(d)** The dUTPase trimer and three active-site regions shown as a stereo ribbon diagram. The individual subunits of the trimer are colored orange, green and purple, with the coil regions colored pale blue except for the C-terminal loop extension, which is colored by subunit. The β-strand interchange among subunits interlocks the trimer. The enzyme active sites are shown by the bound dUMP molecules (pale blue tubes and colored balls: red, oxygen; blue, nitrogen) observed in the enzyme–dUMP crystal structure. (Figure displayed using standard and local modules of the AVS graphics system [AVS Incorporated, Waltham, Massachusetts].)

interactions with Phe5 and Pro15 of the neighboring subunit. Taken together these intersubunit interactions are extensive, with 42% (3500/8300 Å^2) of the solvent-accessible subunit surface buried in the trimer, and major contributions to this buried interface coming from the β8 strand exchange.

Active-site uracil-recognition pockets

The distribution of conserved residues (Fig. 5) on the folded structure and assembled trimer is strikingly asymmetric: most conservation falls predominantly within and around the three clefts between adjacent subunits and in the pocket between β5 and β6 (Fig. 6a) within each subunit. Moreover, mutations that affect enzyme activity cluster to these sequence-conserved and charged clefts, and the dUTPase–nucleotide structures identify them as the substrate-binding sites (Figs 3,4d). In the uncomplexed structure (Fig. 2) these three active-site clefts, and the regions abutting them at the three subunit interfaces, contain charged and polar amino acids and several ordered solvent molecules. The electrostatic potential in these regions is highly complementary for binding the negatively-charged dUTP substrate and positively-charged metal ions (Fig. 6b,c). Within each of these clefts, the uracil inserts

Figure 5



Human dUTPase amino-acid sequence, secondary structure and residue conservation and homology with *E. coli* dUTPase. The human dUTPase amino-acid sequence (top line) is listed along with the sequence of *E. coli* dUTPase (bottom line). Numbering of every tenth amino acid is listed beneath the sequences and is with respect to human dUTPase. Secondary structure elements of human dUTPase are listed above the sequence and labeled as described in the text. Strictly conserved residues among the nine aligned dUTPase sequences (see Materials and methods section) are boxed in pink, with conserved residues in yellow.

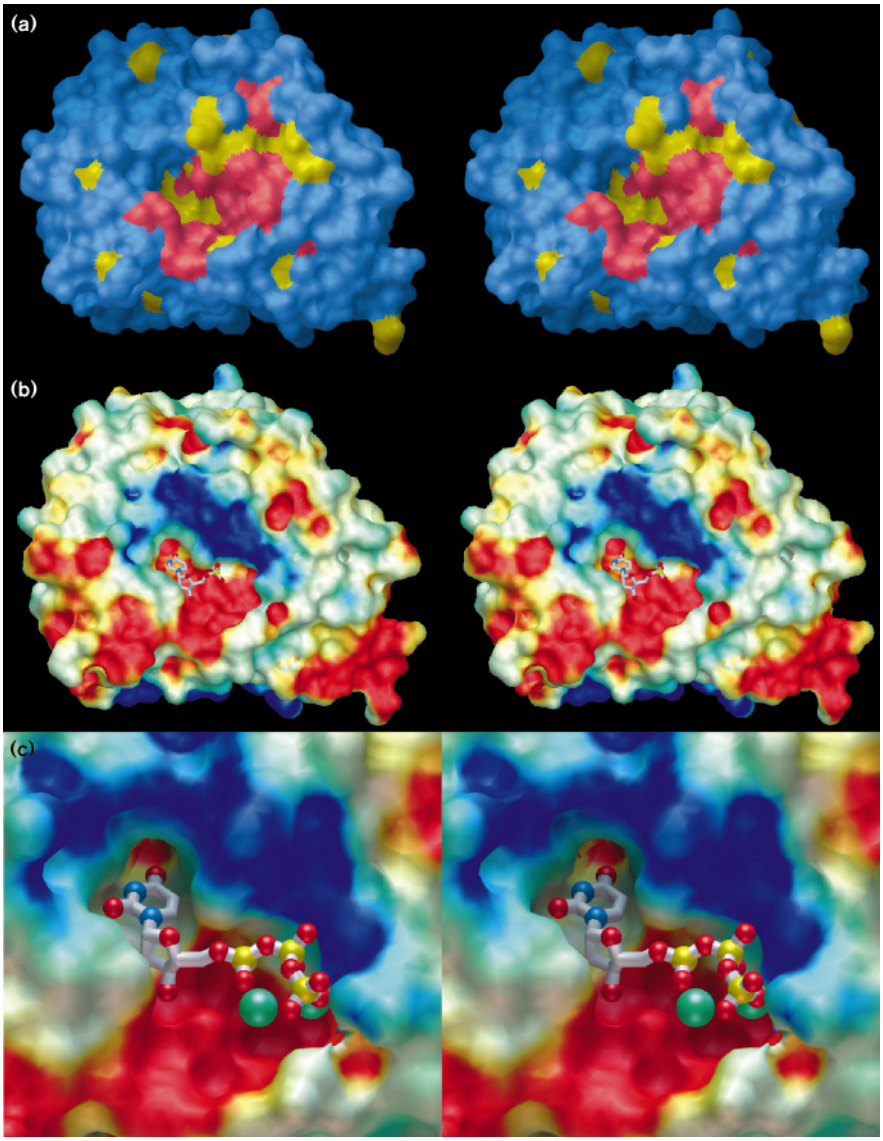
into the dUTPase β sheet by binding in the pocket between β5 and β6 formed by the bulge at β5 Ala75 in each of the three subunits.

Recognition of the uracil deoxyribose nucleoside is accomplished almost entirely by this unique β5–β6 hairpin motif (Fig. 7). There is a hydrogen bond between the amide of β5 Gly74 and the carbonyl of β6 Val89, and β5 is then interrupted by the Ala75 bulge (Fig. 2) allowing the uracil ring insertion between adjacent antiparallel β strands (Fig. 3a). There is a conserved structural water molecule, present in the uncomplexed as well as all three nucleotide-bound complex structures, that forms hydrogen bonds to the β6 Val89 amide and β5 Gly74 carbonyl and also to uracil O4 in the nucleotide-bound structures. The main-chain hydrogen bonds of β6 Val89 to β5 Gly74 via water, and to β5 Gly76 via water and uracil O4, form a β-bulge hydrogen-bonding pattern [28]. Uracil insertion makes two more β-strand hydrogen bonds to β6: uracil N3 to the carbonyl of Gly87 and uracil O2 to Gly87 amide, which complete the mimicry of base-pair hydrogen bonds by the β-hairpin polypeptide backbone. These hydrogen bonds effectively discriminate against cytosine, which cannot donate a hydrogen bond to the Gly87 carbonyl or accept a hydrogen bond from the Gly76 amide. The close approach of Ala75 Cβ (~3.7 Å), Gly74 N (~3.6 Å) and the adjacent subunit Ser63 Cβ (~3.5 Å) to the uracil C5 position excludes dTTP binding, which is consistent with the exquisite base specificity seen in the activity data (Fig. 1).

The human dUTPase active-site pocket within each subunit specifically recognizes a deoxyribose sugar. The nucleoside-specific binding pocket is completed by conserved β5 Ile78, which lies beneath the uracil ring and the C2′ edge of the deoxyribose sugar. Direct β5–β6 hairpin hydrogen bonds occur again at the Asp79 to Tyr82 tight turn. This hydrogen bonding positions the tyrosine aromatic ring to pack with the deoxyribose while sterically blocking ribose binding (tyrosine to sugar *2′* position ~3.5 Å; Figs 7,8), and the conserved Asp79 Oδ1 to hydrogen bond to the deoxyribose O3′ (~2.6 Å) (Fig. 3a–c,8). The tyrosine ring position is also stabilized by hydrogen bonds to the carbonyl of Asn85 (~2.6 Å) and potentially to the uracil O2 atom (~3.8 Å). Thus, hydrogen bonding and steric interactions within the β5–β6 hairpin form the specific recognition pocket for the uracil deoxyribose nucleoside. Equivalent Tyr82 discrimination against ribonucleosides carrying a *2′* hydroxyl would be expected for the tyrosines in *E. coli* [29] and feline immunodeficiency virus [30] dUTPases that have been implicated in catalysis, by biochemical studies. This tyrosine-ring position and hydrogen bond to Asn85 at the end of the β hairpin is unusual as such tyrosine corners are normally found at the end of a Greek-key connection (β arch) in most β-barrel proteins, but essentially nowhere else [31]. These tyrosine corners are thus implicated in β-barrel folding and characteristically lie at the barrel end furthest from the active site. In the dUTPase structure, in contrast, the tyrosine corner is forming part of the deoxyribose specificity pocket. The potential for the Asp79 side chain to interact with dUTP O3′ was noted from sequence similarities between dUTPases and phosphofructokinases [32]. The observed interactions of equivalent residues to dUTPase Ile78 and Asp79 with a deoxyribose sugar also occur in a phosphofructokinase-fructose-1,6-bisphosphate complex [33].

Based on the specificity determinants in the active site of human dUTPase, comparisons can be made with the *E. coli* structure, where the active site was not known [24]. The overall size and shape of the uracil-binding pocket is conserved in the *E. coli* enzyme although the residues that correspond to Gly87 and Val89 in human dUTPase are a methionine and a serine, respectively, in *E. coli* dUTPase (Fig. 5). The human and *E. coli* enzymes differ significantly in the critical area surrounding the uracil C5 position. There is an insertion of an asparagine and a leucine residue in *E. coli* dUTPase (Fig. 5), at the approximate position of Ala75 in the human enzyme, that causes the polypeptide chain to protrude into the channel at the molecular threefold, and

Figure 6



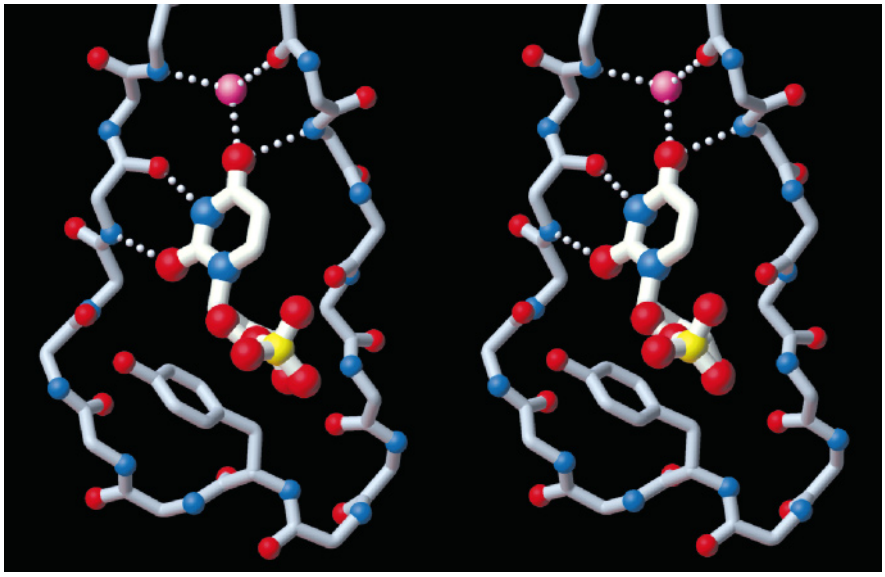
The human dUTPase molecular surface at the active-site regions. (a) The molecular surface viewed to show the active-site region is colored by residue conservation for strictly conserved residues (red), highly conserved residues (gold), and non-conserved residues (blue). Conserved residues cluster in and around the cleft at the subunit interface, and the pocket between the β5–β6 hairpin. The C-terminal tail of the third subunit, which caps the substrate-bound active sites, has been truncated at Asp126 (gold knob, lower right). The glycine-rich flexible tail extends from this position to the conserved active-site region at the subunit interface. (b) Electrostatic potential at the molecular surface, color-coded for positive (blue), negative (red) and near neutral (white) potential, as viewed in (a). The dUMP observed in the enzyme–dUMP complex structure (oxygen, red; nitrogen, blue; phosphate, yellow), binds within the surface pocket. Negative electrostatic potential below the dUMP phosphate group is appropriately positioned to bind the divalent metal ions required for catalysis, whereas the area of positive electrostatic potential above and to the right of the dUMP phosphate binds pyrophosphate in the dUTP-bound structure. The bluish-white knob positioned behind and above the dUMP phosphate (center) is the side chain of α1 Ser63. (c) Details of the dUTPase electrostatic potential and molecular surface interaction with bound dUTP. The dUTP position is taken from the enzyme–dUTP complex structure. The molecular surface reveals the complementary fit of uracil into the β-hairpin motif (deep pocket center left), the close contact from Tyr82 that discriminates against the ribose 20 hydroxyl (center left below pocket), and the negative potential (bottom center red surface) expected to favor metal-ion binding in the presence of the nucleotide phosphates. The exact phosphate positions and metal ions (cyan spheres) are modeled as described in the text. (Figure displayed using standard and local modules of the AVS graphics system.)

contributes to the more hydrophobic nature of this interface in the *E. coli* versus the human enzyme. Atoms equivalent to the Gly74 carbonyl and Ala75 Cβ in human dUTPase are likely to be the Oδ1 and Nδ2 atoms of an asparagine side chain.

Some aspects of these dUTPase interactions resemble those in other nucleotide-binding proteins. dUTPase main-chain hydrogen bonds recognize the uracil O2, N3 and O4 atoms by mimicking base-pair hydrogen bonds (Fig. 7). These hydrogen bonds partly resemble adenine recognition by the catalytic subunit of cAMP-dependent protein kinase (cAPK) [34–36] and the cyclin-dependent kinase 2 (CDK2) [37]. In these kinase complexes, adenine N6 forms a hydrogen bond to a backbone carbonyl and adenine N1 to a backbone amide, which is similar to the uracil O4 and N3 interactions with the amide of Gly76 and carbonyl of Gly87, respectively (Fig. 3a). In contrast to these main-chain interactions to the nucleic-acid base, the three other enzymes that specifically recognize uracil (uracil-DNA glycosylase [2,38], thymidylate synthase [39], and uridylate kinase [40]) all rely on conserved asparagine or glutamine side chains within assembled protein surface pockets to specifically recognize the N3 and O4 atoms of uracil. The dUTPase β-hairpin motif, thus, confers base specificity while allowing variation in amino-acid residues. In addition, dUTPase resembles cAPK and CDK2 in that the base portion of the nucleotide is buried in a hydrophobic pocket, unlike the family of mononucleotide-binding proteins typified by adenylate kinase [41] and the G protein H-*ras*-p21 [42] in which the nucleic-acid base is bound on the surface of the

Figure 7

The novel β-hairpin recognition motif for uracil in dUTPase, with a tyrosine corner motif providing the discrimination against ribose. The bound dUMP (standard atom coloring) and β5–β6 hairpin are from the enzyme–dUMP crystal structure. In each of the three active sites, uracil is recognized entirely by main-chain hydrogen bonds (dotted white lines) to a β-hairpin motif formed by β5–β6. The tightly bound, conserved water molecule is shown as a pink sphere. (Figure displayed using standard and local modules of the AVS graphics system.)

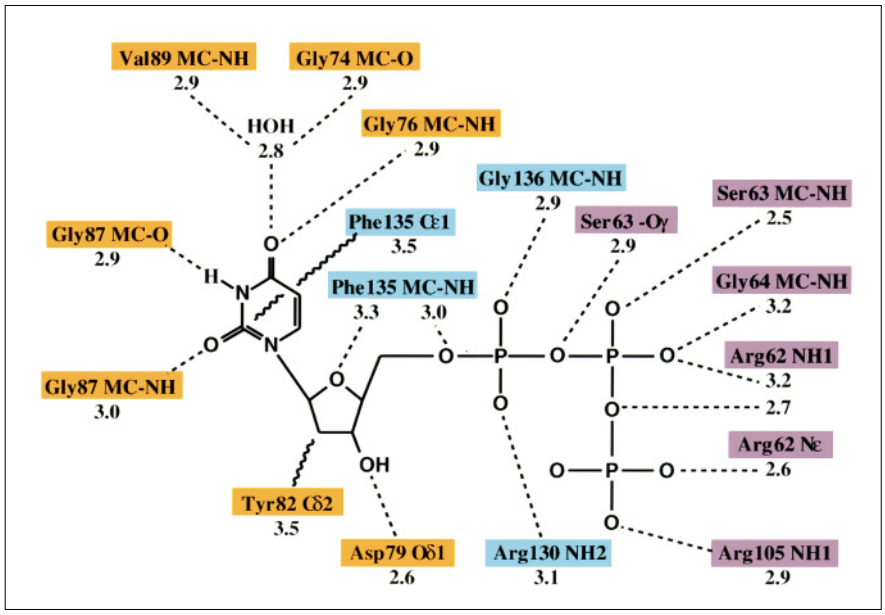


protein. But dUTPase differs significantly from the protein kinases in that the hydrophobic pocket for uracil is formed by insertion into an anti-parallel β hairpin whereas the hydrophobic pocket in cAPK and CDK2 is formed by a β sheet and a loop connecting two domains, with the adenine ring packing against the β sheet and not inserting into it. Thus, the specific recognition of a nucleic-acid base via insertion into an antiparallel β hairpin by dUTPase is likely to represent a new protein structural motif for nucleic-acid recognition, which shares some similar interactions with existing motifs.

This dUTPase structure provides a simple and elegant way for a β hairpin, which is one of the most basic and common elements of protein secondary structure, to recognize nucleic acids via a novel β-strand insertion structural motif. Such interactions could resemble those in the ribosome, where RNA and polypeptide scaffolds may assist in mutual folding to achieve a properly folded catalytic structure. As dUTPase is the enzyme primarily responsible for excluding uracil from DNA, and we find global structural conservation between eukaryotic and bacterial enzymes, it is likely to represent both an

Figure 8

Schematic of dUTPase–substrate interactions showing the structural basis for the exquisite specificity for dUTP that requires all three subunits. Hydrogen bonds (< 3.4 Å) are shown (dashed lines) with the donor–acceptor atom distance beneath the protein atom label. Hydrophobic interactions (wavy lines) are shown only for the key residues Tyr82, which packs against the deoxyribose, and Phe135, which stacks above the bound uracil base. Uracil and deoxyribose are primarily recognized by one subunit (gold rectangles), phosphate groups by the adjacent subunit (purple rectangles), whereas the bound substrate is capped by residues from the C-terminal tail of a third subunit (blue rectangles).



ancient enzyme activity and an early recognition motif. Such a β-strand motif for uracil recognition may thus be representative of the early evolution of interactions between nucleic acid and peptide, when the exclusion of uracil from DNA presumably took place. This relatively simple RNA–peptide association may have been the seed from which higher-order structures eventually evolved. Experimental support for this idea that the β-hairpin motif may have initially recognized RNA nucleosides comes from the recruitment of the tyrosine corner within the β hairpin to discriminate against the ribose *2′* hydroxyl. Such tyrosine corners are otherwise found only in Greek-key (β arch) connections, where they probably stabilize the β arch relative to the hairpin connection [31]. As the tyrosine corner is considered a local sequence motif that strongly favors tertiary folding against a β-hairpin connection, the tyrosine corner within the dUTPase β hairpin would seem to reflect an evolutionary development occurring after the overall folding was otherwise determined.

Interactions with the phosphate groups

Enzyme residues that interact with the phosphate groups come almost exclusively from the subunit adjacent to that which recognizes the uracil and deoxyribose sugar (Fig. 8). The position of the α-phosphate group is conserved in all of the enzyme–nucleotide complex structures and binds near the positive dipole of α1 (Fig. 3a–c and Fig. 4–d): interaction between a helical dipole and a phosphate is often seen in nucleotide-binding proteins [43]. In the dUTPase–dUMP complex, the α phosphate is within direct hydrogen-bonding distance of the α1 Ser63 Oγ atom from the subunit adjacent to that forming the β-hairpin pocket, which binds the deoxyuridine (Figs 3a,8). In the enzyme complexes with dUDP and dUTP, however, the Ser63 Oγ is also within hydrogen-bonding distance of an oxygen from the β phosphate (Fig. 3b,c). The orientation of the β-phosphate group is stabilized by two short hydrogen bonds between separate phosphate oxygens and the backbone amides of Ser63 and Gly64 (Fig. 8), and there is an electrostatic interaction between the β phosphate and the guanidinium of Arg62 (Fig. 3c).

In the dUTPase–dUTP complex, the α and β phosphates bind identically to those in the dUMP and dUDP complexes, but the γ phosphate binds in different conformations in the three crystallographically independent active sites. Although the electron density for the γ-phosphate is less clear than that for the α- and β-phosphates (Fig. 3a–c) in all three enzyme-nucleotide structures, dUTP binding is apparently accompanied by an Arg62 side-chain rotation that maintains an interaction with the β-phosphate. The γ-phosphate occupies the vacated Arg62 side-chain position, allowing an interaction with the guanidinium of the third conserved arginine, Arg105 (Figs 5,8). This γ-phosphate position apparently creates an unfavourable electrostatic interaction with the α-phosphate forming a site that would be highly favourable for a bridging divalent metal ion. Fully productive dUTP binding may, therefore, involve a bridging Mg^{2+} ion bound between the α- and γ-phosphates (Fig. 6c). Such a position for a metal ion has been observed before in the crystal structures of both the murine [34] and porcine heart [36] catalytic subunits of cAPK. With an Mg^{2+} ion bound in this location, conserved Asp79 is too far away (~6 Å) to be a metal ligand, but may instead have a catalytic role. Asp79 Oδ2 does form a hydrogen bond, facilitated through a tightly bound water molecule, with Gln108 Oε1 (Fig. 2). This strictly conserved glutamine residue also forms a hydrogen bond from its Nε2 atom to the Pro61 carbonyl. These considerations suggest that Gln108 may play a structural role and assist catalysis by ensuring the correct water structure in the vicinity of the bound metal ion and scissile phosphoester bond.

Conformational variability of the triphosphate groups of protein-bound nucleotides analogous to those of the dUTPase complexes has been described in several structures of protein kinases in complex with ATP. This variability appears to be linked to the number and positions of divalent metal ions. In the MAP kinase ERK2 ternary complex with Mg^{2+} and ATP, only one metal ion bridges the α- and β-phosphates, and the γ-phosphate of the ATP is disordered [44]. In the G protein H-*ras* p21 complexed with Mg^{2+} and a GTP analog, a single metal ion bridges the β- and γ-phosphates [42], and the triphosphate is in a more extended conformation. In the cAPK complexes, two Mg^{2+} ions bridge the β- and γ-phosphates and α- and γ-phosphates, respectively [34,36]. But in the CDK2 complex structure with Mg^{2+} and ATP the conformation of the β-phosphate is drastically different from that observed in the H-*ras* p21 and cAPK structures and a single Mg^{2+} ion bridges the oxygens of all three of the α-, β- and γ-phosphates [37]. The position for the β-phosphate in the dUTPase complexes with dUDP and dUTP is clearly defined (Fig. 3b,c), and the conformation of the triphosphate is similar to that seen in the cAPK structures, suggesting that two Mg^{2+} ions can be accommodated in the dUTPase nucleotide-binding site.

The major change in dUTPase upon nucleotide binding is the ordering of the C-terminal tails (detailed below), and changes to other residue conformations are not significant. There is, however, an increased ordering of the residues that interact with the bound nucleotides that is reflected in a general decrease in their temperature values, particularly for the β5–β6 distorted antiparallel β hairpin into which the uracil base inserts. The temperature values for these residues are on average 4–6 $Å^2$ lower in the bound versus free structures. Asp79 has a temperature factor decrease of nearly 10 $Å^2$, which is comparable with the average temperature value shift seen for residues Arg62

and Arg105 that interact with the phosphates. Aside from these interactions, with one subunit in each of the three active sites interacting with the uracil and deoxyribose inserting into the β hairpin and the other subunit interacting with the phosphate groups, the remaining critical interactions with bound nucleotides all involve conserved residues in the flexible C-terminal tail of the third subunit of the enzyme (Fig. 8).

Ordering of the flexible C-terminal tails

The glycine-rich C-terminal tails, which are disordered in the uncomplexed structure, become ordered upon ligand binding to each of the enzyme's active sites (Fig. 3a–c). Although the electron density is strongest for the dUTPase–dUDP complex, all of the enzyme–ligand complexes show that both of the two strictly conserved non-glycine residues in this flexible tail, Phe135 and Arg130 (Fig. 5), interact specifically with bound substrate (Fig. 3b,c). Phe135 caps the uracil-recognition pocket by stacking over the uracil ring (Fig. 8). Such a 'Phe-lid' in an enzyme active site has been observed before in the structure of human hypoxanthine-guanine phosphoribosyltransferase in complex with GMP [45], and may hold the uracil in place while excluding water from the active site. The Phe135 uracil interaction may also help ensure the correct Arg130 position. The Arg130 Nε atom forms a hydrogen bond with conserved Asp81 Oδ1 and its guanidinium group lies adjacent to the α-phosphate. Here, it may facilitate the reaction by stabilizing the negative charge that develops on the α-phosphate when the phosphoester bond linking the α- and β-phosphates is cleaved (Fig. 3b). The polypeptide chain between Arg130 and Phe135 makes a sharp turn, with Gly131 possessing main-chain torsion angles that are disallowed for residues with a β carbon. The backbone carbonyl of Gly131 and Oγ atom of Ser132 form hydrogen bonds to the backbone amide and carbonyl, respectively, of Arg83 of the β hairpin that binds the uracil moiety. Residue Gly134 is strictly conserved as the side chain of any other amino acid would clash with the O4′ atom of the deoxyribose sugar. There are also two hydrogen bonds observed from the backbone amides of Phe135 and Gly136 to the O5′ and O2P oxygens of the α-phosphate group (Fig. 8). In the complexes with dUDP and dUTP, the Gly136 amide could also potentially form hydrogen bonds with the β-phosphate group. Although residues Ser137 to Asn141 are not included in the model, the protein-chain direction suggests that these residue side chains and protein backbone may also interact with the phosphate groups of dUTP.

These interactions of a glycine-rich polypeptide segment with bound nucleotides are a relatively common feature of nucleotide-binding proteins [46]. The classic dinucleotide-binding proteins, such as the dehydrogenases that utilize NAD [47] or the oxidoreductases that use FAD [48], have the consensus sequence Gly–X–Gly–X–X–Gly, where X denotes any amino acid. The first two glycines contact the dinucleotide phosphates and are required at these positions owing to their unfavorable main-chain dihedral angles. The other glycine is part of an α helix that forms the characteristic βαβ fold that directs the helix dipole at the phosphates [43]. In the mononucleotide-binding proteins typified by adenylate kinase [49] and H-*ras* p21 [42,50], the consensus sequence for the glycine-rich loop is Gly–X–X–Gly–X–Gly–Lys with the final two glycines again having dihedral angles that are unfavorable for residues with a side chain. In these proteins, the glycine-rich loop forms a 'giant anion hole' [51] that interacts with nucleotide phosphate groups and the lysine side chain forms hydrogen bonds with oxygens of both the β- and γ-phosphates.

The dUTPase glycine-rich tail also has a counterpart in the protein kinases cAPK and CDK2. The glycine-rich loops in cAPK and CDK2 have a consensus Gly–X–Gly–X–X–Gly sequence that is identical to the dinucleotide-binding proteins, but this loop connects two antiparallel β strands without an intervening α helix directing its dipole at the nucleotide phosphates. The protein kinases also have a lysine side chain that interacts with the phosphates, but this lysine is from a separate β strand to those connected by the glycine-rich loop. Superposition of the bound nucleotides from the cAPK and CDK2 structures onto the nucleotides bound to dUTPase (not shown) reveals some similarities between their respective binding sites. Both the kinase loop and the dUTPase tail cover the bound nucleotide almost completely with both approaching the nucleotide from the same side. At the position of Phe130 in dUTPase, which stacks over the uracil, there is a conserved valine in the kinases that packs against the hydrophobic adenine ring of the ATP. The dUTPase glycine-rich tail, thus, appears to have some features in common with the glycine-rich loops in other nucleotide-binding proteins, but its consensus sequence, Arg–Gly–X–Gly–Gly–Phe–Gly, is unlike any described so far. Glycine residues may be conserved for polypeptide segment flexibility, their main-chain dihedral angles (Gly131), steric reasons (Gly134) or directing their backbone amides at the nucleotide phosphates (Gly136) (Fig. 8). The dUTPase nucleotide-binding site thus incorporates features from other nucleotide-binding proteins, including the glycine-rich loop and an α helix whose dipole is directed at the phosphate groups, but the binding-site architecture and sequence motifs are apparently unrelated to previously determined protein structures.

Each of the three human dUTPase nucleotide-binding sites is notable in that it is formed by parts of three separate subunits (Fig. 8). Binding sites between two subunits have been observed before with either a nucleotide [52,53] or a phosphorylated sugar [54] substrate. However, the structure of human dUTPase that we have described is to our knowledge the first documented instance of all

three subunits of a trimeric enzyme coming together to form each of the three active sites with the contribution from each subunit being significant: each subunit supplies residues that are critical to enzyme function and catalysis in all three active sites.

Site-directed mutants, activity and ternary-complex model

To investigate the roles of residues identified from the dUTPase–nucleotide structures as being potentially involved in dUTP binding and catalysis, site-directed mutants at three residue positions were generated and assayed for dUTP and dTTP binding and for enzymatic activity: subunit one β hairpin (Ala75→Gly), subunit two α1 (Ser63→Cys, Ser63→Ala), and subunit three C-terminal tail (Phe135→Ala). In designing these three mutants, care was taken to make conservative substitutions in order to minimize the possibility that the mutation results in an enzyme that is misfolded and, therefore, inactive. For this same reason, residues that had obvious structural roles, such as Gln108, were not selected for mutation. These three site-directed mutants exhibited similar chromatographic behaviour to the wild-type enzyme during the purification steps and also all three mutants formed trimers as judged by native gel electrophoresis (data not shown). The Ala75→Gly substitution was designed to test the role of the distorted β-hairpin in uracil recognition and binding (Fig. 7), whereas the Ser63 substitutions should influence the binding site for the α-phosphate group and its interactions with the helical dipole of α1. The Phe135→Ala mutation examines the role of capping the substrate-bound active site by the C-terminal tail (Fig. 8).

Thus, active-site residues that tested the importance to binding and catalysis of all three subunits were mutated. All three mutations resulted in reduced dUTP binding and enzyme activity (data not shown). In particular, the isosteric substitution of cysteine for Ser63 resulted in a mutant enzyme with greatly reduced activity, as did substitution of Ser63 with alanine. These results suggest a role for Ser63 in properly orienting the α-phosphate and perhaps also in stabilizing the developing negative charge on this group via the α1 helical dipole. The impaired enzyme activity of Ala75→Gly may reflect an inability of this mutant to bind uracil, as a glycine may relieve the distortion in the β5–β6 hairpin resulting in the formation of cross-strand antiparallel β hairpin hydrogen bonds and the collapse of the uracil-recognition pocket (Fig. 7). Similarly, the importance of capping the active site by the C-terminal tail is shown by greatly reduced activity for the substitution of Phe135 by alanine. The stacking of the phenylalanine side chain with the bound uracil thus appears pivotal for ordering the C-terminal tail (Fig. 8). This C-terminal cap may be critical not only for bringing the Arg130 side chain into position to interact with the dUTP α-phosphate, but also for ensuring active-site exclusion from bulk solvent.

Based upon the available structural, biochemical and mutational data for dUTPases, we can begin to elucidate the interactions that take place between the enzyme and bound dUTP substrate, and the potential catalytic roles of conserved enzyme residues. A model for the ternary complex between dUTPase, the dUTP substrate and divalent Mg^{2+} ions was constructed based on the observed interactions in the enzyme–nucleotide complexes and on the strictly conserved enzyme residues likely to have important catalytic roles. The interactions seen in the crystal structure of cAPK in complex with Mg^{2+} and ATP [34–36] were also taken into account. The dUTPase–nucleotide complexes suggest that the phosphate groups are not all in the exact positions they would occupy in a productively bound ternary complex, but rather are reacting to the absence of metal ions by slight shifts to avoid the negatively-charged and probable metal-ion binding region of the enzyme active site (Fig. 6b,c). In the model of the productive complex, the dUTP β-phosphate interacts specifically with Arg62 and the γ-phosphate with Arg105. This orientation places the side chain of the strictly conserved residue Asp26 in position to ligate a Mg^{2+} ion that bridges the oxygens of the β- and γ-phosphates. As seen with cAPK, a second Mg^{2+} ion, bridging the oxygens of the α- and γ-phosphates, may be centrally located in the dUTPase active site. The absolute requirement of dUTPase for divalent metal ions suggests that a metal ion plays a role in catalysis. Although this requirement for metal ions may be purely structural, facilitating a general coalescence of the catalytic enzyme residues around the bound metal ions and phosphate groups, the structurally implicated central location of the second Mg^{2+} ion suggests that it is positioned to facilitate the attack of an activated nucleophile on the α-phosphate.

Although many potential reaction mechanisms can be envisioned, the location and particular environment around Asp79 suggests a role in catalysis. The pK_a of Asp79 may be altered, as the active site is excluded from bulk solvent when capped by the C-terminal tail, so Asp79 may become protonated by stripping a hydrogen from a bound water molecule to generate a hydroxy nucleophile. This nucleophile would be positioned for a direct in-line attack on the α-phosphate, resulting in an inversion of its configuration and generating the resonance-stabilized pyrophosphate leaving group. The charge neutralization of the dUTP β- and γ-phosphates by Arg62 and Arg105 may stabilize the pyrophosphate leaving group and polarize the scissile bond prior to, or in concert with, attack of the nucleophile.

After scissile-bond cleavage, the flexible C-terminal arm could also play a role in ensuring that the enzyme is once again capable of binding a fresh dUTP substrate. As with many other DNA-repair enzymes, dUTPase is faced with the problem that the product of the reaction, in this case dUMP, is also an inhibitor of the enzyme. After bond cleavage, the particular interactions that result in the

packing of the C-terminal arm above the active site are likely destabilized so the arm would once again become less ordered. In this process, the attractive forces of the Phe135 ring and Arg130 guanidinium with the dUMP uracil and phosphate might be sufficient to dislodge the dUMP from the active site thereby promoting product removal and fresh dUTP binding. This model also suggests that there may be an allosteric contribution to productive ternary-complex formation. Mg^{2+}-dUTP binding may be accompanied by slight subunit domain rotations that close down one active site with a concomitant opening up of a distant active site. Thus, although the three dUTPase active sites are composed of identical enzyme residues, they may not be strictly equivalent, and the biologic trimer may have evolved to ensure that at least one active site is always available to bind dUTP from solution.

Biological implications

dUTPase is an essential enzyme that ensures the fidelity of DNA replication, transcription and base-excision repair by hydrolyzing dUTP to dUMP and pyrophosphate. By providing the dUMP for dTTP biosynthesis, dUTPase ensures that the cellular dTTP:dUTP ratio is kept at a high level. High cellular dUTP concentrations lead to repetitive cycles of uracil misincorporation and removal that cause multiple DNA-strand breaks and eventual cell death. This cycle of thymine-less cell death is targeted by chemotherapeutic drugs designed against thymidylate synthase and dihydrofolate reductase and underscores the critical biological role of dUTPase. To further understand the structural basis for dUTPase specificity and activity, and to assist in the design of potential, anti-cancer inhibitors of the enzyme, we have cloned, overexpressed and characterized the activity of human dUTPase. We crystallized this recombinant human enzyme and determined its crystal structure to characterize its fold and trimeric assembly. Crystal structures of the enzyme in complex with dUMP, dUDP and dUTP nucleotides reveal the structural basis for its extreme specificity. This specificity is critical for both DNA and RNA integrity and is essential to cell survival.

Together, these human dUTPase structures establish that the trimeric enzyme has three active sites, which occur at the subunit interfaces. Uracil recognition is achieved by insertion of the uracil ring into a distorted **β** hairpin, forming specific hydrogen bonds with main-chain atoms of the enzyme, mimicking DNA base-pairing and precluding cytosine binding. Thymine- and ribose-sugar binding are prevented by steric clashes with conserved enzyme residues. Interactions with the phosphate groups are primarily with the subunit adjacent to that binding the uracil and deoxyribose moieties, while a glycine-rich, flexible C-terminal tail of the third subunit becomes ordered upon substrate binding and caps the substrate-bound active site. Each productive dUTPase active-site complex is thus formed by elements from all three subunits of the biological trimer.

Besides identifying the active site and basis for dUTP recognition and specificity, these structures reveal a novel motif for interactions between nucleic acid and polypeptide, with potential implications for the evolution of the RNA world. Base exposure in RNA allows for folded, catalytic RNA molecules and has suggested mechanisms for RNA self-replication. The atomic-resolution structures of human dUTPase–nucleotide complexes reveal a simple and elegant way for the unique power of base-pairing recognition to be transferred between nucleic-acid and polypeptide conformation, for one of the most common elements of protein structure, the antiparallel **β** hairpin. Thus, an exposed uracil can guide the formation of complexes between nucleic acid and polypeptide backbone that leave amino-acid side chains available to do chemistry. Furthermore, this work suggests mechanisms for primordial RNA–polypeptide interactions thought to be ancestral to modern DNA replication. These uracil–**β**-hairpin interactions, which mimic base pairing, therefore suggest an obvious way for peptides to become early coenzymes in an RNA world. The importance of this interaction is supported by the association of some DNA-repair enzymes with the ribosome [55] and their early evolutionary development, implied by the conservation between bacterial- and eukaryotic-enzyme structures. Given evolution by alteration of existing macromolecules, the detailed structural example of a specific uracil–polypeptide backbone interaction provides a plausible mechanism for the development of RNA–peptide interactions antecedent to contemporary ribosomes. In fact, the unusual association of a tyrosine corner, specific to **β**-barrel Greek key connections, with the dUTPase **β** hairpin that recognizes uracil, along with the involvement of this tyrosine in selecting deoxyribose over ribose, suggests that the tyrosine and consequently deoxyribose specificity developed after the **β**-hairpin folding was otherwise determined. Overall, the results from the four high-resolution structures presented here provide novel and detailed insights into the exquisite specificity of dUTPase for dUTP, the structural and catalytic implications of the biological trimer, and possible mechanisms for phosphoester-bond cleavage. As the structure of a viral dUTPase has been solved [56] and a complete description of this viral enzyme can be expected soon, comparisons among the human, *E. coli*, and viral enzymes should enhance understanding of dUTPase action and aid design of species-specific enzyme inhibitors.

Materials and methods

Expression and purification of human dUTPases

Human dUTPase was expressed in *E. coli* using plasmid pEM409. This vector consists of a 2.9 kilobase BamH1–EcoR1 fragment of pBluescript

KS+ (Stratagene), a 110 bp BglII–NdeI fragment of pET11-a (Novagen) containing a phage T7 promoter linked to the vector by a BamH1/BglII fusion, and the human dUTPase open reading frame inserted between the NdeI and EcoR1 sites. The human dUTPase was modified by PCR to contain six histidine codons immediately following the first ATG codon. The PCR primers were designed such that in the amplification product an NdeI site encompassed the start codon, and the stop codon was flanked by an EcoR1 site. This facilitated cloning of the modified cDNA on an Nde1–EcoR1 fragment into the expression vector. The resulting clone was sequenced to verify that no extraneous mutations were introduced during the PCR.

E. coli strain BL21(DE3) was transformed with pEM409 and a single amp^r colony was then used to inoculate 5 l of 2× yeast tryptone medium containing 50 μg–ml^{-1} ampicillin. The culture was grown at 32°C to an OD_{600} of 1.0 at which point dUTPase expression was induced by addition of IPTG to give a final concentration of 0.2 mM. Cells were incubated for an additional 90 min and harvested by centrifugation. Cell pellets were resuspended in a minimal volume of sonication buffer (50 mM sodium phosphate pH 8.0 and 300 mM NaCl) and lysozyme added to give a 1 mg ml^{-1} solution. This suspension was incubated at ambient temperature for 30 min and phenylmethylsulfonyl fluoride (PMSF) added to a final concentration of 100 μM. At this point cells were disrupted by three freeze thaw cycles using dry ice/ethanol followed by three 40 s bursts of a Braun sonicator fitted with a microtip while the cell suspension was cooled in ice. All subsequent steps were carried out at 4° C. Cell debris was removed by centrifugation at 10 000×g for 30 mins and the supernatant loaded onto a 10 ml column of Ni^{2+} agarose pre-equilibrated with sonication buffer. The column was washed with sonication buffer until the OD_{280} of the column eluate was less than 0.01, at which point any remaining non-specifically bound proteins were removed by washing with wash buffer (50 mM sodium phosphate pH 6.0, 300 mM NaCl, 10 % glycerol) until the OD_{280} had again reached 0.01. Histidine fusion protein was eluted with a 30 ml gradient of 0–0.5 M histidine in wash buffer. Elution was monitored at OD_{280} and eluted protein immediately loaded onto a 5×120 cm S300 Sephadex gel-filtration column that had been pre-equilibrated with TBS. Protein was eluted with TBS at a flow rate of 1.3 ml min^{-1} and protein-containing fractions collected according to OD_{280}. The purity of the final fraction was assessed by SDS PAGE and found to be greater than 99 % by scanning densitometry.

Enzyme reactions contained 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 10 mM DTT in a final volume of 50 μl. Reaction mixtures contained 0.02 μg of purified enzyme with 200 μM substrate dUTP. When the other, non-substrate nucleotide triphosphates were tested, 1.0 μg of enzyme was present. Assays were performed at 37°C and initiated by addition of enzyme. Reactions were terminated by boiling for 3 min followed by centrifugation (11 000×g for 10 min) to remove insoluble material. The supernatant containing the reaction products was analyzed by HPLC on a Waters 8PSAX strong anion exchange column using 0.5 M KH_2PO_4 pH 3.35 (flow rate 2.0 ml min^{-1}) as the mobile phase. Nucleotide metabolites in the assay were identified by comparison of retention times with those of individual nucleotide standards. The dUTPase activity was calculated by monitoring the rate of decrease in the area of the dUTP peak over fixed periods of time.

Site-directed mutagenesis

Site-directed mutants of human dUTPase were generated using a modified polymerase chain reaction (PCR) technique. Mutant Phe135→Ala was produced by PCR amplification using a mutant 3′ PCR primer. The other mutants (Ser63→Cys, Ser63→Ala, Ala75→Gly, Tyr82→Ala and Tyr82→Phe) were produced by amplifying the dUTPase gene in two separate portions, the first portion extending from the DNA encoding the enzyme's N terminus and terminating in a mutagenic oligonucleotide, the second portion extending from the 3′ end of the mutagenic oligonucleotide to the C terminus of the enzyme. Mutants were produced by ligating these products, generating a single stretch of DNA containing an internal point mutation. Ligation between identical products was prevented by including restriction enzymes Nde1 and EcoR1 in the ligation reaction. These enzymes cut at the restriction sites in the 5′ oligo of the first product and in the 3′ oligo of the second product. Amplification with PFU (a proof reading polymerase) ensured that the termini of the products to be ligated were blunt ended. The ligation products were further amplified to facilitate subcloning into pBSK+ cut with Hinc2. All mutations were confirmed by sequencing in vector pBSK before subcloning into the Nde1/EcoR1 sites of pEM409. Mutant enzymes were expressed in pEM409 and purified as described for the native enzyme.

Amino acid sequence conservation

The deduced amino acid sequences for nine different dUTPase enzymes were aligned using a local multiple sequence alignment program (developed by CM Bruns) that performs a global alignment sequentially based on an empirical scoring matrix. The sequences were input into the program and aligned in the following order: human [20], *E. coli* [57], *Saccharomyces cerevisiae* [12], a herpesvirus [58], tomato [59], vaccinia virus [60], avian adenovirus [61], a parapoxvirus [62], and a swine poxvirus [63].

Crystallization and data collection

Crystals of human dUTPase were grown at room temperature by the hanging drop vapor diffusion technique. Equal volumes (3 μl) of the protein solution (20 mg ml^{-1}), in 20 mM N-(2-hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid) (HEPES), pH 7.5, and 100 mM NaCl, were mixed with reservoir solutions containing 1.0–1.2 M sodium citrate (Sigma), pH 8.0, and 100 mM imidazole/malate buffer, pH 7.5 to 8.5. The crystals grow in two to three days, and contain three dUTPase subunits per crystallographic asymmetric unit. These crystals have a V_M [64] of 2.2 Å^3 Da^{-1}, which is within the range observed for other protein crystals.

X-ray diffraction data for dUTPase and for the complexes were collected at room temperature using a 30 cm MAR research imaging plate and a typical oscillation width of 1.0° per frame. Complete data sets were collected from a single crystal of typical dimensions 0.4 mm× 0.3 mm×0.1 mm at a wavelength of 1.54 Å using a Siemens rotating anode X-ray generator operating at 50 kV and 100 mA and an exposure time of 15 mins per frame. Data for the enzyme–nucleotide complexes were collected from single crystals that had been soaked in reservoir solution containing 50 mM of either dUMP, dUDP or dUTP for 18 h. All data were processed using the programs DENZO and SCALEPACK [65]. Attempts to soak divalent metal ions into the crystals were precluded by the high concentration of sodium citrate in the mother liquor and the fact that the crystals are not stable in solutions lacking citrate. The high citrate concentration did, however, allow us to collect data on the dUTPase–dUTP complex. The substrate dUTP would otherwise have been cleaved and released.

Molecular replacement and phasing

Molecular replacement calculations were performed with a local version of the AMoRe suite of programs [66] by using as a model the crystallographic trimer of *E. coli* dUTPase (Brookhaven Protein Data Bank entry code 1DUP). Rotation function and translation search calculations were performed with all data (5 237 reflections, 96 % complete) in the resolution range 10.0–3.5 Å. The search results gave three top hits each separated from the other by a rotation of $\gamma = 120°$ and identical translation vectors corresponding to the three hits expected for a symmetric trimer. The correlation coefficient for these three peaks was 20.9 and the R value 0.508 ($R = \Sigma|F_{obs} - F_{calc}|/\Sigma|F_{obs}|$), while the corresponding values for the highest background peak were 12.0 and 0.537, respectively.

Model building and refinement

The correctly oriented and positioned dUTPase model was subjected to rigid-body refinement of the trimer and then of the individual monomers, followed by Powell conjugate-gradient energy minimization using X-PLOR version 3.1 [67], with an overall temperature value of

20.0 Å^2. The R value dropped to 0.41 for 8406 reflections (92% complete) in the resolution range 10.0–3.0 Å. The value of R_{free} for 5% of reflections against which the model was not refined [68] was 0.49. This model was used to calculate phases weighted to reduce bias using the σ_A program [69], and to calculate σ_A-weighted $2F_{obs}-F_{calc}$ and $F_{obs}-F_{calc}$ electron-density maps with the XtalView suite of programs [70], which were visually inspected with the interactive molecular graphics program XFIT [71]. In regions where the polypeptide backbone appeared correct, the amino-acid sequence of human dUTPase [20] was placed in the electron density. In regions where the trace of the polypeptide backbone was obviously in error, residues were deleted. Residues that appeared to be correctly positioned but whose exact sequence identity could not be established were replaced with alanines. This partial model, containing a total of 106 residues for each monomer, was refined further in X-PLOR yielding values of $R = 0.36$ and $R_{free} = 0.40$ for data from 10.0–3.0 Å resolution. Inspection of σ_A-weighted $2F_{obs}-F_{calc}$ and $F_{obs}-F_{calc}$ electron-density maps calculated with this partial model clearly showed the trace of the polypeptide backbone for the regions of the model that had been omitted and the correct amino-acid sequence was fitted to these maps. Further refinement of this revised model was performed in two cycles increasing the resolution limit in 0.2 Å increments, and electron density maps calculated to 2.6 Å resolution. Throughout these early stages of model building and refinement only one of the three noncrystallographic subunits was inspected and fitted to the electron-density maps. This subunit was then superimposed on the remaining two subunits and refinement carried out in X-PLOR using noncrystallographic symmetry (NCS) restraints. Subsequent rounds of refinement and map fitting were performed without NCS restraints, and each subunit visually inspected and fitted to the electron density individually.

With the amino-acid sequence for human dUTPase fitted for each subunit, refinement commenced using rounds of stereochemically-restrained simulated annealing and positional refinement with X-PLOR, followed by refinement of the individual atomic temperature values. The data to 2.0 Å resolution were included in the refinement in 0.1 Å increments, interspersed with manual inspection and rebuilding of the structure. Areas of the structure where the electron density was difficult to interpret, or where the main-chain torsional angles were in unfavorable areas of the Ramachandran plot [72], were inspected in simulated-annealed omit maps in which approximately ten residues were omitted from the phase calculation. Ordered solvent molecules were manually fit to $F_{obs}-F_{calc}$ electron-density maps and included in the refinement at higher resolutions. Following refinement with X-PLOR, the model was then refined using the TNT suite of programs [73] including all of the low-resolution data. Throughout the course of model building and refinement, progress was monitored by the consistent decrease in the values of both the conventional crystallographic R value and the free R value. Refinement of the human dUTPase–dUMP, dUTPase–dUDP and dUTPase–dUTP structures were done in an analogous manner to that of the uncomplexed structure.

Surface potential was calculated with DELPHI (Biosym Technologies, Inc.) with AMBER partial charges, an ionic strength of 145 mM, and solute and solvent dielectric constants of 2.0 and 80.0, respectively. Solvent-accessible surfaces were calculated with a probe radius of 1.4 Å.

Accession numbers

Coordinates are being deposited at the Brookhaven Protein Data Bank.

Note added in proof

After submission of this paper, the crystal structure of *E. coli* dUTPase complexed with dUDP was published [74], which confirms the comparisons proposed here for the active sites in the human and bacterial enzymes. However, in the reported *E. coli* dUTPase–dUDP complex, the enzyme–phosphate interactions are mainly through water molecules: the C-terminal tails, which cap each of the three active sites in the human dUTPase–ligand complexes, were not seen in the electron density.

Acknowledgements

Correspondence should be addressed to JAT. The authors thank AS Arvai, KJ Balch, RJ Bishop, CM Bruns, BR Crane, DS Daniels, ED Getzoff, JE Johnson, G Joyce, TJ Macke, ME Pique, CD Putnam, G Slupphaug and CD Stout for helpful advice and discussions. This work was supported by the National Institutes of Health grant GM46312. JMH is supported by an Australian Commonwealth AIDS Research post-graduate fellowship. CDM is the recipient of a Special Fellowship from the Leukemia Society of America.

References

1. Lindahl, T. (1994). Instability and decay of the primary structure of DNA. *Nature* 362, 709–715.
2. Mol, C.D., *et al.*, & Tainer, J.A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* 80, 869–878.
3. Tye, B.K., Nyman, P.O., Lehman, I.R., Hochhauser, S. & Weiss, B. (1977). Transient accumulation of Okazaki fragments as a result of uracil incorporation into nascent DNA. *Proc. Natl. Acad. Sci. USA* 74, 154–157.
4. El-Hajj, H.H., Wang, L. & Weiss, B. (1992). Multiple mutant of *Escherichia coli* synthesizing virtually thymineless DNA during limited growth. *J. Bacteriol.* 174, 4450–4456.
5. Pu, W.T. & Struhl, K. (1992). Uracil interference, a rapid and general method for defining protein-DNA interactions involving the 5-methyl group of thymines: the GCN4-DNA complex. *Nucleic Acids Res.* 20, 771–775.
6. Ivarie, R. (1987). Thymine methyls and DNA-protein interactions. *Nucleic Acids Res.* 15, 9975–9983.
7. Ingraham, H.A., Dickey, L. & Goulian, M. (1986). DNA fragmentation and cytotoxicity from increased cellular deoxyuridylate. *Biochemistry* 25, 3225–3230.
8. Richards, R.G., Sowers, L.C., Laszlo, J. & Sedwick, W.D. (1984). The occurrence and consequences of deoxyuridine in DNA. *Adv. Enzyme Regul.* 22, 157–185.
9. Barclay, B.J., Kunz, B.A., Little, J.G., & Haynes, R.H. (1982). Genetic and biochemical consequences of thymidylate stress. *Can. J. Biochem.* 60, 172–194.
10. Goulian, M., Bleile, B. & Tseng, B.Y. (1980). The effect of methotrexate on levels of dUTP in animal cells. *J. Biol. Chem.* 255, 10630–10637.
11. El-Hajj, H., Zhang, H. & Weiss, B. (1988). Lethality of a dut (deoxyuridine triphosphatase) mutation in *Escherichia coli*. *J. Bacteriol.* 170, 1069–1075.
12. Gadsden, M.H., McIntosh, E.M., Game, J.C., Wilson, P.J. & Haynes, R.H. (1993). dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. *EMBO J.* 12, 4425–4431.
13. Nation, M.D., Guzder, S.N., Giroir, L.E. & Deutsch, W.A. (1989). Control of *Drosophila* deoxyuridine triphosphatase. *Biochem. J.* 259, 593–596.
14. Pyles, R.B., Sawtell, N.M. & Thompson, R.L. (1992). Herpes simplex type I dUTPase mutants are attenuated for neurovirulence, neuroinvasiveness and reactivation from latency. *J. Virol.* 66, 6706–6713.
15. Steagall, W.K., Robek, M.D., Perry, S.T., Fuller, F.J. & Payne, S.L. (1995). Incorporation of uracil into viral DNA correlates with reduced replication of EIAV in macrophages. *Virol.* 210, 302–313.
16. Bouhamdan, M., *et al.*, & Sire, J. (1996). Human immunodeficiency virus type 1 vpr protein binds to the uracil DNA glycosylase DNA repair enzyme. *J. Virol.* 70, 697–704.
17. Ladner, R.D., McNulty, D.E., Carr, S.A., Roberts, G.D. & Caradonna, S.J. (1996). Characterization of distinct nuclear and mitochondrial forms of human deoxyuridine triphosphate nucleotidohydrolase. *J. Biol. Chem.* 271, 7745–7751.
18. Ladner, R.D., Carr, S.A., Huddleston, M.J., McNulty, D.E. & Caradonna, S.J. (1996). Identification of a consensus cyclin-dependent kinase phosphorylation site unique to the nuclear form of human deoxyuridine triphosphate nucleotidohydrolase. *J. Biol. Chem.* 271, 7752–7757.
19. Beck, W.R., Wright, G.E., Nusbaum, N.J., Chang, J.D. & Isselbacher, E.M. (1986). Enhancement of methotrexate toxicity by uracil analogues that inhibit deoxyuridine triphosphate nucleotidohydrolase (dUTPase) activity. *Adv. Exp. Med. Biol. B* 195, 97–104.
20. McIntosh, E.M., Ager, D.D., Gadsden, M.H. & Haynes, R.H. (1992). Human dUTP pyrophosphatase: cDNA sequence and potential biological importance of the enzyme. *Proc. Natl. Acad. Sci. USA* 89, 8020–8024.
21. Williams, M.V. (1986). Effects of mercury (II) compounds on the activity of dUTPases from various sources. *Mol. Pharm.* 29, 288–292.
22. Williams, M.V., Holliday, J. & Glaser, R. (1985). Induction of a deoxyuridine triphosphate nucleotidohydrolase activity in Epstein-Barr virus-infected cells. *Virology* 142, 326–333.

23. Williams, M.V. (1988). Herpes simplex virus-induced dUTPase: a target site for antiviral chemotherapy. *Virol.* 166, 262–264.
24. Cedergren-Zeppezauer, E.S., Larsson, G., Nyman, P.O., Dauter, Z. & Wilson, K.S. (1992). Crystal structure of a dUTPase. *Nature* 355, 740–743.
25. Williams, M.V. & Cheng, Y. (1979). Human deoxyuridine triphisphate nucleotidohydrolase. *J. Biol. Chem.* 254, 2897–2901.
26. Yamaguchi, M. & Sakurai, T. (1992). Reversible effect of calcium-binding protein regucalcin on the Ca^{2+}-induced inhibition of deoxyuridine 5′-triphosphatase activity in rat liver cytosol. *Mol. Cell. Biochem.* 110, 25–29.
27. Richardson, J.S. & Richardson, D.C. (1990). Principles and patterns of protein conformation. In *Prediction of Protein Structure and the Principles of Protein Conformation*. (Fasman, G.D., ed.), pp. 2–98, Plenum Press, NY, USA.
28. Richardson, J.S., Getzoff, E.D. & Richardson, D.C. (1978). The beta bulge: a common small unit of non-repetitive protein structure. *Proc. Natl. Acad. Sci. USA* 75, 2574–2578.
29. Vertessy, B.G., Zalud, P., Nyman, P.O. & Zeppezauer, M. (1994). Identification of tyrosine as a functional residue in the active site of *Escherichia coli* dUTPase. *Biochem. Biophys. Acta* 1205, 146–150.
30. Wagaman, P.C., Hasselkus-Light, C.S., Henson, M., Lerner, D.L., Phillips, T.R. & Elder, J.H. (1993). Molecular cloning and characterization of deoxyuridine triphosphatase from feline immunodeficiency virus (FIV). *Virology* 196, 451–457.
31. Hemmingsen, J.M., Gernert, K.M., Richardson, J.S. & Richardson, D.C. (1994). The tyrosine corner: a feature of most Greek key β-barrel proteins. *Prot. Sci.* 3, 1927–1937.
32. McGeoch, D.J. (1990). Protein sequence comparisons show that the `pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Res.* 18, 4105–4110.
33. Shirakihara, Y. & Evans, P.R. (1989). Crystal structure of the complex of phosphofructokinase from *Escherichia coli* with its reaction products. *J. Mol. Biol.* 204, 973–994.
34. Zheng, J., *et al.*, & Sowadski, J.M. (1993). Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* 32, 2154–2161.
35. Zheng, J., *et al.*, & Sowadski, J.M. (1993). 2.2 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Cryst. D* 49, 362–365.
36. Bossemeyer, D., Engh, R.A., Kinzel, V., Postingl, H. & Huber, R. (1993). Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart deduced from the 2.0 Å structure of the complex with Mn^{2+} adenylyl imido-phosphate and inhibitor peptide PKI(5–24). *EMBO J.* 12, 849–859.
37. De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O. & Kim, S.-H. (1993). Crystal structure of cyclin-dependent kinase 2. *Nature* 363, 595–602.
38. Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* 373, 487–493.
39. Montfort, W.R., *et al.*, & Stroud, R.M. (1990). Structure, multiple site binding, and segmental accommodation in thymidylate synthase on binding dUMP and an anti-folate. *Biochemistry* 29, 6964–6977.
40. Mueller-Dieckmann, H.-J. & Schulz, G.E. (1995). Substrate specificity and assembly of the catalytic center derived from two structures of ligated uridylate kinase. *J. Mol. Biol.* 246, 522–530.
41. Diederichs, K. & Schulz, G.E. (1991). The refined structure of the complex between adenylate kinase from beef heart mitochondrial matrix and its substrate AMP at 1.85 Å Resolution. *J. Mol. Biol.* 217, 541–549.
42. Pai, E.F., Kabsch, W., Krengel, U., Holmes, K.C., John, J. & Wittinghofer, A. (1989). Structure of the guanine-nucleotide-binding domain of the H-*ras* oncogene product p21 in the triphosphate conformation. *Nature* 341, 209–214.
43. Wierenga, R.K., De Maeyer, M.C.H. & Hol, W.G.J. (1985). Interaction of pyrophosphate moieties with α-helices in dinucleotide binding proteins. *Biochemistry* 24, 1346–1357.
44. Zhang, F., Strand, A., Robbins, D., Cobb, M.H. & Goldsmith, E.J. (1994). Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. *Nature* 367, 704–710.
45. Eads, J.C., Scapin, G., Xu, Y., Grubmeyer, C. & Sacchettini, J.C. (1994). The crystal structure of human hypoxanthine-guanine phosphoribosyltransferase with bound GMP. *Cell* 78, 325–334.
46. Schulz, G.E. (1992). Binding of nucleotides by proteins. *Curr. Opin Struct. Biol.* 2, 61–67.
47. Rossman, M.G., Liljas, A., Branden, C.-I. & Banaszak, L.J. (1975). Evolutionary and structural relationships among dehydrogenases. In *The Enzymes, Vol 11*. (Boyer, P., ed.), pp. 61–102, Academic Press, NY, USA.
48. Schulz, G.E. (1980). Gene duplication in glutathione reductase. *J. Mol. Biol.* 138, 335–347.
49. Walker, J.E., Saraste, M., Runswick, M.J. & Gay, N.J. (1982). Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 8, 945–951.
50. Pai, E.F., Petsko, G.A., Goody, R.S., Kabsch, W. & Wittinghofer, A. (1990). Refined crystal structure of the triphosphate conformation of H *ras* p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *EMBO J.* 9, 2351–2359.
51. Dreusicke, D. & Schulz, G.E. (1986). The glycine-rich loop of adenylate kinase forms a giant anion hole. *FEBS Letts* 208, 301–304.
52. Morera, S., *et al.*, & Janin, J. (1994). Adenosine 5′-diphosphate binding and the active site of nucleoside diphosphate kinase. *Biochemistry* 33, 459–467.
53. Georgiadis, M.M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J.J. & Rees, D.C. (1992). Crystallographic structure of the nitrogenase iron protein from *azotobacter vinelandii*. *Science* 257, 1653–1659.
54. Evans, P.R., Farrants, G.W. & Hudson, P.J. (1981). Phosphofructokinase: structure and control. *Phil. Trans. Roy. Soc. Ser. B* 293, 53–62.
55. Wool, I.G. (1996). Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* 21, 164–165.
56. Prasad, G.S., *et al.*, & Stout, C.D. (1996). Crystal structure of FIV dUTP pyrophosphatase. In *Retroviruses, Cold Spring Harbor Symposium, May 21–26*, 217.
57. Lundberg, L.G., Thoressen, H.-O., Karlstrom, O.H. & Nyman, P.O. (1983). Nucleotide sequence of the structural gene for dUTPase of *Escherichia coli* K-12. *EMBO J.* 2, 967–971.
58. Davison, A.J. (1992). Channel catfish virus: a new type of herpesvirus. *Virology* 186, 9–14.
59. Pri-Hadash, A., Harevan, D. & Lifschitz, E. (1992). A meristem-related gene from tomato encodes a dUTPase: analysis of expression in vegetative and floral meristems. *Plant Cell.* 4, 149–159.
60. Slabaugh, M.B. & Roseman, N.A. (1989). Retroviral protease-like gene in the vaccinia virus genome. *Proc. Natl. Acad. Sci. USA* 86, 4152–4155.
61. Akopian, T.A., Kaverina, E.N., Naroditskii, B.S. & Tikhonenko, T.I. (1992). Analysis of the nucleotide sequence of a fragment (92–100 %) of the CELO avian adenovirus genome. *Mol. Gen. Mikrobiol. Virusol.* 11–12, 19–23.
62. Mercer, A.A., Fraser, K.M., Stockwell, P.A. & Robinson, A.J. (1989). A homologue of retroviral pseudoproteases in the parapoxvirus, orf virus. *Virology* 172, 665–668.
63. Massung, R.F., Jayarama, V. & Moyer, R.W. (1993). DNA sequence analysis of conserved and unique regions of swinepox virus: identification of genetic elements supporting phenotypic observations including a novel G protein-coupled receptor homologue. *Virology* 197, 511–528.
64. Matthews, B.W. (1968). Solvent content in protein crystals. *J. Mol. Biol.* 33, 491–497.
65. Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend.* (Sawyer, L., Issac, N. & Bailey, S., ed), pp. 56–62, SERC Daresbury Laboratory, Warrington, UK.
66. Navaza, J. (1992). AMoRe: A new package for molecular replacement. In *Proceedings of the CCP4 Study Weekend*. (Dodson, E.J., Grower, S. & Wolf, W., eds), pp. 87–91, SERC, Daresbury Laboratory, Warrington, UK.
67. Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* 235, 458–460.
68. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475.
69. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* 42, 140–149.
70. McRee, D.E. (1992). XtalView: a visual protein crystallographic software system for X11/XView. *J. Mol. Graphics* 10, 44–47.
71. McRee, D.E. (1993). *Practical Protein Crystallography*. Academic Press, San Diego, USA.
72. Ramachandran, G.N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. In *Advances in Protein Chemistry, Vol 23*. (Anfinsen, C.B.Jr., Anson, M.L., Edsall, J.T. & Richards, F.M., eds), pp 283–438, Academic Press, NY, USA.
73. Tronrud, D.E., Ten Eyck, L.F. & Matthews, B.W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr. A* 43, 489–501.
74. Larsson, G., Svensson, L.A. & Nyman, P.O. (1996). Crystal structure of the *Escherichia coli* dUTPase in complex with a substrate analogue (dUDP). *Nat. Struct. Biol.* 3, 532–538.